A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis

Paramita Ray
Dinabandhu Andrews Institute of Technology & Management, Kolkata, India, and
Amlan Chakrabarti
A.K. Choudhury School of Information Technology, University of Calcutta, Kolkata, India

Received 26 September 2018
Revised 31 January 2019
Accepted 22 February 2019

Abstract

Social networks have changed the communication patterns significantly. Information available from different social networking sites can be well utilized for the analysis of users opinion. Hence, the organizations would benefit through the development of a platform, which can analyze public sentiments in the social media about their products and services to provide a value addition in their business process. Over the last few years, deep learning is very popular in the areas of image classification, speech recognition, etc. However, research on the use of deep learning method in sentiment analysis is limited. It has been observed that in some cases the existing machine learning methods for sentiment analysis fail to extract some implicit aspects and might not be very useful. Therefore, we propose a deep learning approach for aspect extraction from text and analysis of users sentiment corresponding to the aspect. A seven layer deep convolutional neural network (CNN) is used to tag each aspect in the opinionated sentences. We have combined deep learning approach with a set of rule-based approach to improve the performance of aspect extraction method as well as sentiment scoring method. We have also tried to improve the existing rule-based approach of aspect extraction by aspect categorization with a predefined set of aspect categories using clustering method and compared our proposed method with some of the state-of-the-art methods. It has been observed that the overall accuracy of our proposed method is 0.87 while that of the other state-of-the-art methods like modified rule-based method and CNN are 0.75 and 0.80 respectively. The overall accuracy of our proposed method shows an increment of 7–12% from that of the state-of-the-art methods.

Keywords Opinion mining, CNN, Sentiment analysis, Text mining, POS tagging
Paper type Original Article

Publishers note: The publisher wishes to inform readers that the article "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis" was originally published by the previous publisher of *Applied Computing and Informatics* and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Ray, P., Chakrabarti, A. (2022), "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis", *Applied Computing and Informatics*. Vol. 18 No. 1/2, pp. 163-178. The original publication date for this paper was 04/03/2019.

Applied Computing and Informatics
Vol. 18 No. 1/2, 2022
pp. 163-178
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1016/j.aci.2019.02.002

1. Introduction

Nowadays, various social networking sites like Twitter, Facebook, Linkedin, etc. have become a popular platform for exchanging opinions, and thus providing feedback regarding specific products, organization, services, movies, individuals, political events, topics [34,37], etc., which can help to improve the product quality and services [1,3,9] of an organization. Aspect-level opinion mining establishes a relation between different aspects of an item and its polarity. An aspect of a product means an attribute or feature of a product. For sentiment analysis, the identification of aspects is a very important issue. There are two types of aspect, explicit aspect, and implicit aspect. For example, in the sentence, "The resolution of the phone is really nice and the phone is affordable", the "resolution" is an aspect of the phone and there is a positive opinion for the phone. In the above example, "resolution" is explicitly mentioned in the text, but "affordable" is an implicit aspect. Explicit aspects are well classified while implicit aspects are not easily classified. So it is difficult to extract implicit aspect. Most of the previous works for aspect level sentiment analysis have used SVM based algorithms [34], conditional random fields (CRFs) [4,7] or some rule-based approaches using natural language processing tools. But all of these methods have some limitations. CRF needs huge number of features to work properly and the rule-based method depends on the grammatical accuracy of sentences and it can only identify explicit aspects but fails to extract implicit aspects. Besides this, sometimes incorrect aspects have been tagged by using parts-of-speech (POS) tagger [10,12] as parts-of-speech tagger considers all the noun or noun phrases as aspect terms. But all nouns are not always relevant aspect terms. Here CNN is used which have much fewer connections and parameters. Every word in the sentences is not labeled in this method and we only need to label the whole sentences. So for large data, we can easily train the network. ReLU (Rectified Linear Units) is the most commonly used activation function in CNN. ReLU is linear (identity) for all positive values, and zero for all negative values. It converges faster and classifies data more easily. In this paper, we have tried to overcome all the limitations of existing methods by using CNN (nonlinear supervised classifier) [15,18,25]. An innovative technique is used to identify the aspects by applying POS Tagging, dependency parsing using CoreNLP [19,21,23] followed by hierarchical clustering, [2,5] which reduces the number of incorrect aspects and extracted aspects are categorized with some predefined set of aspects. Then the improved existing method (CoreNLP + Rule-based) is compared with the CNN based approach of aspect extraction. But it has been observed that CNN classifier sometime fails to identify some valid aspect terms. So a rule-based approach has been introduced which is combined with CNN to further improve the performance of aspect extraction method. In addition to this, we have also improved the sentiment scoring method by introducing seven-point scaling. Most of the previous works used three levels of sentiment classification. But sometimes opinion word is associated with some strong and weak adverb which modify the sentiment score. Here, the users sentiment has been classified into seven groups (Almost Positive, Positive, Very Positive, Almost Negative, Negative, Very Negative, and Neutral). In this paper, product reviews from a popular social networking site called Twitter, SST-1 (Stanford Sentiment Treebank) [50] for the movie review and SemEval Task 4 [45] for Restaurant Review has been collected. Accurate aspect extraction and polarity detection method help for recommendation systems, product quality, and service improvement. It also allows the customer to identify which features are important and which not on the basis of feedback. The rest of the paper is organized as follows, in Section 2, we have presented related work, and Section 3 describes the background on sentiment analysis. Section 4 describes the details of the experimental setup. Section 5 contains results and analysis and Section 6 briefs the concluding part of the paper.

2. Related Work

An introduction to the field of sentiment analysis was found at Pang and Lee's [17,38] article in 2008. Various techniques and methods with both practical and theoretical considerations

were covered by their article. These techniques were used to analyze reviews for movies and products. In 2004, Kim and Hovy [37] and more recently, Bhayani and Huang (2009) [36] (Wilson et al. [29], 2005; Agarwal et al., 2009) [28] performed sentiment analysis on Twitter and Tweets were classified in terms of negative or positive sentiment. Hu and Liu [39] first introduced the concept of aspect extraction from opinions and later this method has been modified by Popescu and Etzioni [40] and by Blair-Goldensohn et al. [41]. A language model was introduced by Popescu and Etzioni. They assumed that product class is known in advance and introduced an algorithm to detect a noun or noun phrase as a product feature. They measured point-wise mutual information between the noun phrase and the product class. Scaffidi et al. [42] used a language model to identify product features. It was assumed in their method that product reviews contain more numbers of product features than in a general natural language text. But it was found that their method has low precision value and extracted aspects are affected by noise. Aspect extraction method for a product was improved by introducing semi-supervised models by Wang et al. [43]. They proposed a model to extract aspects based on the use of seeding aspects. In this method, they used seed words to identify topics of specific interest to a user and extract aspects from the review. Recent approach based on CNN [11,22,27] has also achieved significant improvement in performance over state-of-the-art methods in many traditional NLP tasks [6]. It has been used in different NLP areas such as information retrieval and related classification. A simple network including one-layer convolution and with a max pooling layer has been proposed by Kim et al. [31] which performed sentiment classification successfully. Johnson et al. introduced a bag-of-words model to represent text instead of low-dimensional word vectors, which are extremely effective for text categorization [18]. Here more than 150,000 microblog postings were analyzed containing branding comments, sentiments, and opinions. They investigated the overall structure of these microblog postings [24,27], the types of expressions, and the movement in positive or negative sentiment. They compared automated methods of classifying sentiment in these microblogs with manual coding. Collobert et al. [44] used part-of-speech tagging, chunking, and named entity recognition tasks using a multi-task sequence labeler. Xiaodong Liu et al. [23] proposed a multi-task deep neural network to classify query and search website ranking. Qiongxia Huang et al. [6] in 2017 designed a model with existing CNN, LSTM, CNN-LSTM (Implement of one-layer LSTM directly stacked on one-layer CNN) and SVM (support vector machine). Here, we have introduced a combinational method of CNN and Rule-Based approach which identify aspect terms more accurately than the other state-of-the-art methods.

3. Sentiment Analysis Methodology Background:

3.1 Different Levels of Sentiment Analysis

There are three different levels of sentiment analysis have been proposed.

3.1.1 Document Level: In Document Level sentiment analysis, it is analyzed whether the document expresses a positive or negative sentiment.

3.1.2 Sentence Level: In Sentence Level sentiment analysis, the document is broken into some sentences and each sentence is treated as a single entity and analyzed at a time.

3.1.3 Aspect Level: In Aspect Level, the main task is to extract aspect terms of the product and then customer feedbacks are analyzed on the basis of the extracted aspects.

3.2 Parts of Speech (POS) Tagging

Parts of Speech (POS) tagging is a form of annotating text and each word is a tag with Parts of Speech [28,33]. Tokens are marked with their corresponding word by the POS Tagger [25,26]. Part-of-Speech tags are assigned to character strings. Each sentence can be categorized into a group of determiners, verbs, nouns, etc.

3.3 Dependency Parsing

The grammatical structure of a sentence and the relationships between "Main" words and the word which modify those main words can be obtained through a dependency parser [14,16,20]. Here, we use dependency parser [5,15] for aspect extraction and finding their dependency relation with opinion words. For example, "The phone has a good camera", here "camera" is an aspect and "good" is an opinion word. We can analyze the structure of the sentence "This phone has a good camera" in the following way Figure 1.

Here amod: adjectival modifier, det: determiner, dobj: direct object and nsubj: nominal subject. In the above example, "Camera" and "good" has amod relation. So from the Standford rule(rule 1) "camera" is an aspect of phone and "good" is opinion word.

3.4 Cluster Analysis

Cluster analysis is required in text mining for making a group of objects. It consists of different methods and algorithms to group objects of similar kinds into respective categories. Hierarchical Clustering is used here. This method uses the dissimilarities (similarities) or distances between objects while forming the clusters.

After POS tagging and dependency parsing, lots of aspects are collected. To increase accuracy, aspects are categorized with the predefined set of aspects using hierarchical clustering.

3.5 Convolutional Neural Network (CNN) for Text Classification

CNN is comprised of one or more convolutional layers [8,26,40], which are responsible for major breakthroughs in image classification. More recently, CNN is also applied to problems in Natural Language Processing (NLP) like information retrieval and relation classification, sentiment analysis [8,9,13], spam detection or topic categorization. Sentences or documents that are the input of most NLP tasks can be represented as a matrix where each row represents one token. A token may be a word or a character. The convolutional layers can be represented as the weighted sum of the word vectors with respect to the shared weight matrix. The largest value is selected in the max pooling layer. The behaviour of the CNN is strongly influenced by the magnitude of the word, and all the word vectors from Google word2vec are normalized to one. The max-pooling layer can be increased or decreased by uniform scaling (Up or Down) of word vectors. Each CNN contains a word embedding layer, a convolutional and pooling layer, and a fully-connected layer.

3.5.1 Word Embedding. Word embedding is a method where words or phrases from the vocabulary are mapped to vectors of real numbers. All the words in the input sentence are encoded as word vector. Let the sentence length $l \in R$ and the vocabulary size $D \in R$. The embedding matrix of k-dimensional word vectors is $W^1 \in R^{k|D|}$ and the i^{th} word in a sentence is transformed into a k-dimensional vector w_i by matrix-vector product:

$$w_i = W^1 x_i \tag{1}$$

Here x_i is the one-hot representation for the i-th word.

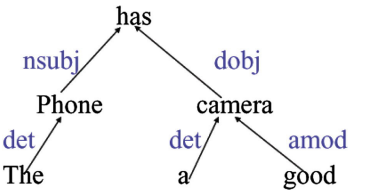

Figure 1.
Example of Dependency Parsing.

3.5.2 Convolution. The convolution operations are applied on the top of the vectors [1] which are generated after encoding the input sentence to produce new features. In convolution operation, a filter $u \in R^{hk}$ is applied to a window of $h = 2r + 1$ and a feature f_i is produced from a window of words $w_{i-r:i+r}$ by

$$f_i = g(w_{i-r:i+r}.u) \tag{2}$$

where g is a non-linear activation function (Relu) and the filter is applied to all possible windows of the input sentence to generate a feature map.

$$f = [f_1, f_2, \ldots f_l] \tag{3}$$

3.5.3 Pooling. In this layer, max-over-time pooling is applied to each of the feature maps that are generated in convolutional layers.

$$\widehat{f} = max(f_1, f_2, \ldots f_l) \tag{4}$$

The maximum element in each feature map is taken by Max-over-time pooling and a fixed-sized feature vector $v_i \in R^{m_i}$ for the i-th task [30]. In this model, one feature is extracted from one filter. The model uses multiple filters (with varying window sizes) to obtain multiple features. In this layer, the feature with the highest value is extracted for each filter. The idea is to capture the most important feature with the highest value.

3.5.4 Dropout Regularization. Deep neural networks suffer from overfitting [32,33] due to the high number of parameters that need to be learned. So dropout regularization is added, which will randomly disable a fraction of neurons in the layer (set to 50% here) to ensure that the model does not overfit. This prevents neurons from co-adapting and forces them to learn individually useful features.

3.5.5 Fully-connection. The features form the dropout layer are passed to a fully connected layer.

$$G = \alpha(w * \widehat{f} + b) \tag{5}$$

where α is the rectified linear (ReLU) activation function. W is the weight matrix, and b is the bias. Bias is added to neural networks to help the network to learn patterns.

3.5.6 Softmax layer. Finally the output of the previous layer is passed to a fully connected softmax layer. It returns the classes with the largest probability.

$$\widehat{y} = \underset{j}{arg\max} P(y = j|x, w, a) = \underset{j}{arg\max}\left(e^{xw_j + a_j} \Big/ \sum_k {}^{=1^k} x e^{xw_k + a_j}\right) \tag{6}$$

where w_j denotes the weights vector of class j and a_j is the bias of class j. The probability forms a discrete probability distribution. The softmax layer returns the classification result, and then the model parameters are updated by the back-propagation algorithm according to the actual classification label of the training data. The above process is operated repeatedly and then the model training is complete.

4. Material and Methods

4.1 Data Collection

In this step, electronics product reviews from a popular social networking site called Twitter, movie reviews from SST-1(Stanford Sentiment Treebank) [50] and Restaurant reviews from SemEval Task 4 [45] has been collected. More than 500,000 reviews of different categories of products are used for analysis Table 1.

Data Source	Category of Products	Reviews	Total
Twitter	**Electronics Products**		250000
	Redmi Note 4	180500	
	Laptop(Dell)	40000	
	Camera(Nikon)	29500	
SST-1 (Stanford Sentiment Treebank) [50]	**Movie Review**		150000
SemEval Task 4 [45]	**Restaurant Review**		100000
Total Reviews			500,000

Table 1. List of source and total number of reviews.

4.2 Data Pre-processing Method

The pre-processing method is used to clean data and convert data in the proper format for further analysis. The commonly known pre-processing methods are as follows: -.

(i) Remove URL link: URL links do not carry much information regarding the sentiment of the tweet. So links are removed from tweets. (ii) Remove numbers: Generally, numbers have no use for measuring sentiment and are removed from the tweets in order to refine the tweet content. (iii) Convert acronyms: Acronyms are ill-formed words and are common in tweets. So acronyms are replaced by the original words through acronym dictionary. (iv) Words in tweets that contain repeated letters have been converted to their original English form. Words with repeated letters, e.g. "coooool" are replaced by "cool". (v) Unnecessary white spaces and tabs are removed. (vi) All tweets are converted to lower case. All these steps are performed using R software with utils package.

4.3 Part-of-Speech(POS) Tagging

Generally, aspects are nouns or noun phrases and opinion words are adjectives. Nouns and noun phrases can be obtained by a part-of-speech (POS) tagger. Basically, six basic parts of speech (noun, verb, adjective, adverb, preposition, conjunction) [35] have been tagged from the reviews to identify aspects and its polarity score. POS tagging step is used in both the methods mentioned below. Stanford CoreNLP is used for POS tagging.

4.4 Modification of existing method by the implementation of some rule-based approach

4.4.1 Dependency Parsing. After part-of-speech tagging, a set of aspects and corresponding opinion words are extracted using some syntactic relations and set of rules. CoreNLP is used for this step.

4.4.2 Aspect Category Determination. After dependency parsing, lots of aspects and opinion words are collected. Collected aspects are compared with the predefined set of aspects to reduce the number of incorrect aspects. Hierarchy clustering method and principal of component analysis (PCA Method) (Figure 2) are used to determine aspect category. For example, in Figure 2 "camera", "image" and "picture" aspects belongs to the same category.

4.5 Aspect Extraction with Convolutional Neural Network (CNN) and some rule-based approach

4.5.1 Word Embeddings. Here, Skip-Gram Model (created by Mikolov et al.) has been implemented for word embedding. This model is a very efficient predictive model for word embeddings from the text. In this model, there is one-dimensional integer vector of the target

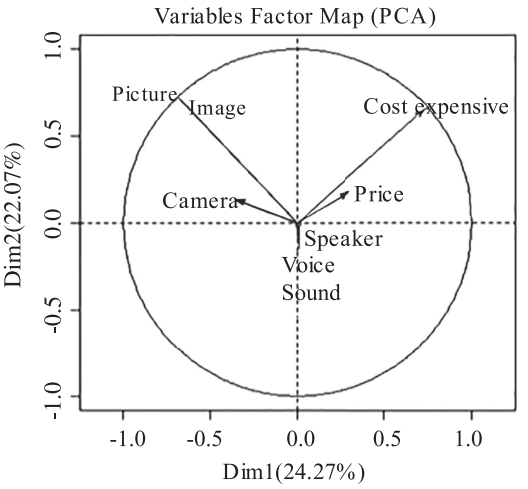

Figure 2.
Aspect categorization using PCA method.

word tokens and one-dimensional integer vector of sampled context word tokens has given as input. If the sampled word really appears in the context, then the prediction is 1 otherwise 0. From a set of sentences (also called the corpus), this Skip-Gram model loops [44] on the words of each sentence and try to find the use of the current word and to predict its neighbours (its context). In some cases, it uses contexts to predict the current word. The upper limit of the number of words in each context can be determined by a parameter called "window size".

4.5.2 Proposed Network Architecture. The architectural overview of our proposed system [Figure 3] is as follows. The input to the system is a set of reviews for a particular product collected from the different data sources. The proposed network contains aspect mappers and a CNN sentiment classifier. Our proposed network contains one input layer, one word embedding layer, two convolutional, two max-pool layers, and a fully-connected layer with softmax output [34]. Word embedding layer encodes each word in the input sentence as a word vector. After Word embedding, the convolution operations are applied on top of these vectors to produce new features. The input layer was 75 x 350, where maximum 75 numbers of words in a sentence can be used with 350 the dimensionality of the word embedding. The first convolution layer has filter size 2 and second convolution layer has filter size 3. The stride in each convolution layer is 1, as we wanted to tag each word. Though the filter size for two convolution layers is different, So the dimensionality of two convolution layers are 2×350 and 3×350 respectively. There is a max-pooling layer followed by each convolution layer and the pool size is 2 in the max-pool layers. The output of each convolution layer has been computed using a non-linear function. In our proposed network, hyperbolic tangent is used. 100 feature maps with filter size 2 in the first convolution layer and 50 feature maps with filter size 3 in second convolution layer. Then all the vectors concatenated together to form a single feature vector (The most important feature with the highest value).

It has been observed that the proposed network architecture works well in our datasets. Increasing layer size, window size or pool size does not improve the performance or accuracy.

4.6 Rule-Based Approaches

In this paper, a set of rules has been combined with the deep learning concept to extract aspects and detect polarity. Here some Stanford Dependencies [50] Rules have been used to

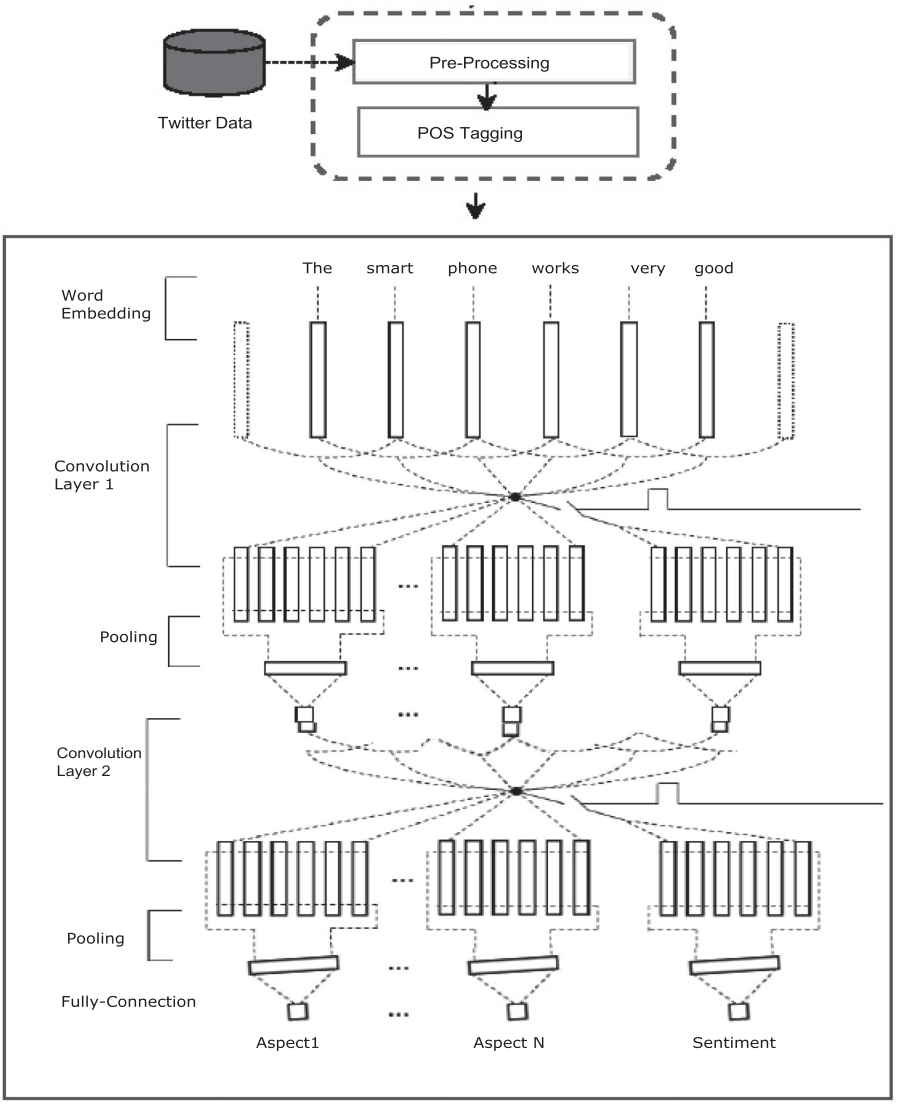

Figure 3.
Architectural overview of our proposed system.

extract aspect. SenticNet 3 [48] has been used as a concept-level opinion lexicon. The following rules are listed below.

4.6.1 Rule-Based Aspect Extraction Approach. Rule 1: **IF depends (amod, A, O)∧ pos (A, nn) ∧opinion word(O), THEN aspect(A).** Here, depends(amod, A, O) shows a dependency relation amod between A and O, pos(A, nn) shows that A is a singular noun, opinion word(O) means that O is an opinion word, aspect(A) means that A is an aspect. It means if there is an amod: adjectival modifier relation between A and O, O is opinion word and A is a noun, then A is identified as an aspect.

Rule 2: **IF depends (conj, Ai, Aj) ∧pos(Ai, nn) ∧aspect(Aj), THEN aspect(Ai).** The above rule states that if there is a conj dependency relation between Ai and Aj and Ai is a noun, then using one aspect Ai we can also extract another aspect Aj. For example, "This phone has a great screen and battery". Here, "screen" aspect has been extracted by the previous rule (Rule 1), and the above rule (Rule 2) can be used to extract "battery" as an aspect. Because "screen" and "battery" has the conj dependency relation.

Rule 3: **If verb n has a direct object and a noun p, then label p as an aspect.**

Rule 4: **If a term is labeled as an aspect by the previous rules and there is a noun-noun compound relationship with another word, then composed of both of them and marked as an aspect.** For example, "Battery Life" "Battery" or "Life" is marked as an aspect, then the whole expression is labeled as an aspect.

4.6.2 Rule-Based Sentiment Evaluation Approach. After part-of-speech tagging, adverbs and adjectives are collected to improve the scoring method and classify emotions with a seven-point scale (Almost Positive, Positive, Very Positive, Almost Negative, Negative, Very Negative and Neutral) [Table 2]. This is obtained by the following scoring method. Here Sentiwordnet [48] is used for polarity detection. Sentiwordnet is the result of the automatic annotation of all the synsets of WORDNET according to the notions of "positive", "negative", and "neutral". Three numerical scores Pos(s), Neg(s), and Obj(s) are associated with each synset. After pre-processing of data, each data token has been parsed with the help of part-of-speech (POS) tagger. POS tagger assigns a tag [9] to each token and then the word is passed in SentiWordnet to check the score as well as the polarity of that particular word.

Sentiment Scoring using Adjectives

- If the aspect is evaluated by a single adjective or opinion word [47], then it returns a sentiment score between −1 and +1, may be "Almost Negative" or "Almost Positive".
- If the aspect is evaluated by double adjectives or opinion words then it returns sentiment score between −2 and +2 may be "Negative" or "Positive".

Sentiment Scoring using Adverb and Adjective

If there is a single adverb present before an adjective, then the sentiment score of the aspect is changed. We can classify adverbs into two categories [3,4,6].

(I) *Strong Intensifying Adverbs:* Adverbs such as exceedingly, extremely, immensely, very and so on, which has a strong effect [47] on sentiment score are called Strong Intensifying Adverbs.

- If $score(adj) > 0$ and $adv \in STRONG$, then $score(adv, adj) > score(adj)$.
- If $score(adj) < 0$ and $adv \in STRONG$, then $score(adv, adj) < score(adj)$.

If a sentence contains double adverbs and if both the adverb is strong intensifying adverbs, then the score will be as follows.

Sentence	Opinion
1. The Phone has good screen.	Almost Positive
2. The Phone has good and attractive screen.	Positive
3. Sound quality is very good.	Very Positive
4. Sound quality is bad.	Almost Negative
5. Sound quality poor and horrible.	Negative
6. Sound quality is extremely bad.	Very Negative
7. I have checked the speaker of the phone.	Neutral

Table 2. Sentiment classification using Adjectives and Adverbs.

- If $score(adj) > 0$ and $adv \in STRONG$, then $score(adv, adv, adj) > score(adv, adj) > score(adj)$.
- If $score(adj) < 0$ and $adv \in STRONG$, then $score(adv, adv, adj) < score(adv, adj) < score(adj)$.

(II) *Weak Intensifying Adverbs:* Like barely, scarcely, weakly, slightly etc. which have little effect on scoring are called Weak Intensifying Adverbs.

- If $score(adj) > 0$ or $score(adj) < 0$ and $adv \in WEAK$, then $score(adv, adj) = score(adj)$.

4.6.3 Handling Negation Word. Negation word converts positive opinion to negative or negative to positive by using special words. The common approach of negation handling is to reverse the polarity of the aspect.

- If $score(adj) > 0$ and Negation word Present, then $score(adj) < 0$.
- If $score(adj) < 0$ and Negation word Present, then $score(adj) > 0$.

Scores of each aspect are then calculated using the following equations:

$$Score(ai) = \sum_{i=1}^{n}((score(adj) + score(adj, adj) + score(adv, adj) + score(adv, adv, adj)) \tag{7}$$

Here ai is the ith aspect and n is the number of aspect terms. If Score $(ai) > 0$ then "Positive", If Score$(ai) < 0$ then "Negative", Otherwise Neutral.

4.7 Training set and Testing set Making

Here, we have used the SemEval datasets 4 [45] for training and testing set and each review is tagged with aspect and polarity related to that aspect. Total 6,086 Sentences (Laptop reviews-3045 and Restaurant reviews-3,041) are annotated for training. The frequency of various aspects in the sentences of train data is shown in Table 3. To predict sentiment we have used both the data for positive and negative, training and testing and we are going to use different machine learning model for the data set. Here, both the rule-based and CNN based classifier are applied to the text; then all terms marked by any of these two classifiers are reported as aspect terms.

4.7.1 Initializing CNN. In our experiment, we have used the mini-batch (size = 12) stochastic gradient descent with momentum set to 0.8 and learning rate = 0.001. Initialize Weights with the small random value.

Categories	Aspects	Frequency	Total train data
Restaurant			3041
	Food	1200	
	Service	241	
	Price	150	
	Ambience	400	
	Miscellaneous	1050	
Laptop			3045
	Battary-Life	341	
	Cost	400	
	Keyboards	654	
	Miscellaneous	1650	

Table 3.
Frequency of aspect categories in train data of SemEval-2014 Task 4 dataset [46].

4.8 Package Used

Packages and other details that are used in this analysis are listed below: -

- 'R' software has been used as the computational environment and some of the packages that have been used are 'readr', 'stringr', 'keras', 'reticulate', 'ggplot2', 'NLP', 'Tm', 'tensorflow' etc.
- Natural Language Processing (NLP) libraries such as the Stanford CoreNLP for POS tagging, Lemmatization, Dependency Parsing, and sentiment analysis are used here. It was introduced by Socher et al. (2013) and it was developed at the University of Stanford to predict the sentiment of movie reviews.
- For Aspect category detection, hierarchical clustering and PCA method have been used. 'Factoextra', 'Cluster', 'FactoMineR' packages are used for this purpose.
- SentiWordNet: The SentiWordNet is a lexical resource, is associated with two numerical scores ranging from 0 to 1 that indicate Pos(s) and Neg(s) describing how positive or negative terms contained in the synsets. Here it is used to identify the polarity of the extracted aspects.
- For Negation Handling list of the following words (Table 4) are used.

5. Results & Analysis

In our proposed method, after part-of-speech tagging and dependency parsing, 360000 numbers of aspects along with customers opinion related to the aspect have been collected from reviews. Aspects are categorized using hierarchical clustering and principal component analysis (PCA Methods) [Figure 2] to reduce incorrect aspects and the sentiment score of the reviews has been analyzed with a seven-point scale. But this method is not suitable to extract implicit aspects. So we have used the CNN to extract more accurate aspects. Word embedding has been used as features for the network. Word embedding in deep learning concept performs better than randomized features where each word vector initialized randomly and when these features are combined with part-of-speech(POS) features, accuracy level has been improved.

Table 5 shows that how the POS feature is important for aspect extraction and improving accuracy. It has been observed that both precision and recall value increases when the POS

can not	shouldn't	doesn't	didn't
don't	hadn't	hasn't	haven't
Couldn't	nor	Without	hardly
wasn't	wouldn't	weren't	neither

Table 4. List of negation words.

Domain	Features	Precision	Recall	F-Score
Cellphone	WE	81.64%	72.15%	79.8%
Cellphone	WE + POS	85.24%	75.4%	82.54%
Camera	WE	72.9%	78.87%	77.30
Camera	WE + POS	76.29%	82.8%	80.5%
Laptop	WE	78.9%	83.23%	80.25%
Laptop	WE + POS	82.25%	85.45%	81.24%
Restaurant	WE	83.56%	86.8%	85.45%
Restaurant	WE + POS	85.67%	88.20%	89.34%

Table 5. Impact of the POS feature over word embedding.

feature is used with word embedding. This method gives better accuracy on restaurant domain reviews and also improved F-score.

Table 6 shows that for restaurant domain only 59% of accurate aspect terms can be obtained by one of the existing methods. So only parts-of-speech tagging is not sufficient for aspect extraction and it produces incorrect aspects in some cases. It is one of the reasons for lower accuracy. After using CNN classifier, accuracy increases to 67% but when we used some rule-based approach with the combination of CNN approach, accuracy level increases to 74.4%. CNN also sometimes fails to identify valid aspects. In that case, the rule-based approach provides a better result. As shown in Table 6, CNN suffered from low recall value in all the domains, i.e., it missed some valid aspect terms. Rule-based analysis of the reviews helped to overcome some of these limitations.

Table 7 shows the comparison of sentiment score and the overall accuracy of all the methods in laptop domain. Sentiment scores of users are classified into seven-point scale like "Positive score" is categorized into 'Almost Positive', 'Positive' and 'Very Positive'. Similarly 'Negative score' is classified into 'Almost Negative', 'Negative' and 'Very Negative'. Otherwise 'Neutral'. It has been found that the overall accuracy of the (CoreNLP + Rule-based), CNN and our proposed method (CNN + Rule-based) is 0.75, 0.80 and 0.87 respectively. So our proposed method performs better than other existing methods.

Table 8 shows the comparison of the state of the art method [49] with our proposed method. It is found from the experiment that some methods only detect one term aspect and

Table 6. Comparison of (**CoreNLP + Rule-Based**), **CNN** and (**CNN + Rule-Based**) method.

Domain	Classifiers	% of Accurate Aspects	Precision	Recall	F-Score
Cellphone	CoreNLP + Rule-based	65.3%	72.4%	75.55%	74.86%
Cellphone	CNN	71.3%	75.68%	85.15%	80.56%
Cellphone	CNN + Rule-based	75%	79.24%	88.4%	82.34%
Camera	CoreNLP + Rule-based	59.8%	73.6%	79.57%	75.50%
Camera	CNN	68.7%	76.6%	88.87%	78.50%
Camera	CNN + Rule-based	72.4%	78.79%	89.9%	80.5%
Laptop	CoreNLP + Rule-based	64.8%	73.9%	81.53%	79.45%
Laptop	CNN	71.4%	76.9%	85.23%	82.35%
Laptop	CNN + Rule-based	77.4%	79.25%	88.45%	83.24%
Restaurant	CoreNLP + Rule-based	59.4%	74.46%	80.8%	79.55%
Restaurant	CNN	67.4%	77.56%	84.8%	81.45%
Restaurant	CNN + Rule-based	74.4%	79.67%	86.20%	83.34%
Movie Review	CoreNLP + Rule-based	63.7%	74.26%	78.8%	75.55%
Movie Review	CNN	69.4%	75.36%	79.8%	78.45%
Movie Review	CNN + Rule-based	75.6%	78.67%	82.20%	80.34%

Table 7. comparison of list of sentiment score and over all accuracy among existing methods and proposed method.

Item	Methods	Positive Score(%)			Negative Score(%)			Neutral Score(%)	Accuracy
		Almost Pos	Pos	Very Pos	Almost Neg	Neg	Very Neg		
Laptop (Dell)	CoreNlp + Rule-Based	42.10	1.05	.10	11.57	.21	.04	44.90	0.75
Laptop (Dell)	CNN	45.28	3.00	.50	4.25	.15	.3	45	0.80
Laptop (Dell)	CNN + Rule-Based	52	5.05	.68	4.57	.24	.5	65.60	0.87

fails to detect aspect phrase. But our proposed rule based approach helps to identify aspect phrases accurately by using Rule 4. So precision and recall value of our method is greater than that of the other methods.

Figure 4 shows an overall accuracy of our proposed method (CNN + Rule-Based) and other two existing methods(CNN and CoreNLP + Rule-Based).

Figure 5 shows the comparison among the proposed method and other two existing methods for laptop and camera domain. Here recall value and precision value of the three methods are compared and it has been observed that our proposed method has greater precision and recall value than other methods.

5.1 Significancy test using Paired T-test

A paired t-test has been performed to check how our proposed method statistically significant. It has been observed from Table 9 that the p-value in paired t-test is.01 which is

Domain	Algorithm	Precision	Recall
Nikon Camera Data Set	Hu and Liu [39]	69.00%	82.00%
	Popescu and Etzioni [40]	86.00%	80.00%
	Dependency propagation [49] method	81.00%	84.00%
	Proposed Method(CNN + Rule-Based)	88.6	90.5

Table 8. Comparison of proposed method with the state of the art method on Nikon Camera dataset.

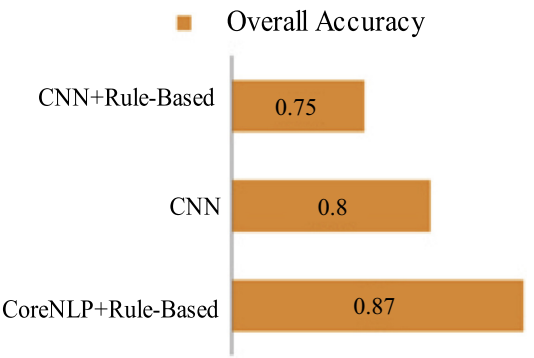

Figure 4. Comparison of the overall accuracy of CoreNLP + Rule-based, CNN and CNN + Rule-based approach on the dataset of laptop domain.

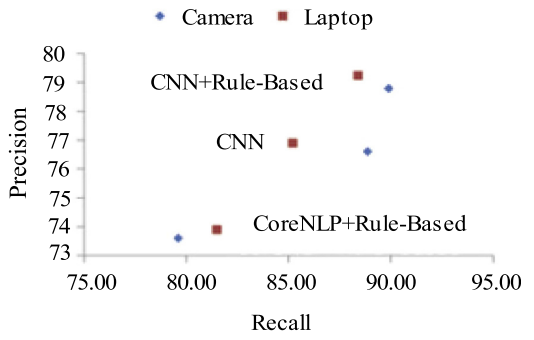

Figure 5. Comparison of the performance of CoreNLP + Rule-based, CNN and CNN + Rule-based approach on the dataset of camera and laptop.

Table 9.
Paired T-Test for significant test.

	Over all Accuracy of Proposed Method	Over all Accuracy of Existing Method (CoreNLP + Rule-based)
Mean	0.87	0.78
Variance	0.0001	0.0007
Observations	3	3
Pearson Correlation	−0.188982237	
Hypothesized Mean Difference	0	
df	2	
t Stat	5.773502692	
P(T<=t) one-tail	0.014357069	
P(T<=t) two-tail	0.028714138	
t Critical two-tail	4.30265273	

less than.05. It indicates that the improvement of our proposed method is statistically significant at the confidence level of 95%.

6. Conclusion

In this paper, a mixed approach of deep learning method and the rule-based method has been introduced for aspect level sentiment analysis by extracting and measuring the aspect level sentiments. On the one hand, we have used machine learning techniques, POS tagging, dependency parsing, etc. to identify the aspects and opinion of user related to the aspect. On the other hand, a seven-layer specific deep CNN architecture has been developed that contains the input layer, consisting of word embedding features for each word in the sentence, two convolution layers, each of them is followed by max-pooling layer, fully connected layer, and the output layer. A rule-based concept is also introduced to improve the performance of aspect extraction. In comparison with the existing methods, the proposed technique (CNN + Rule-Based) brings better classification accuracy for both the positive and negative classes.

References

[1] Samir Tartir, Ibrahim Abdul-Nabi, Semantic sentiment analysis in arabic social media, J. King Saud University – Comput. Inform. Sci. 29 (2) (April 2017) 229– 233.

[2] Jaspreet Singh, Gurvinder Singh, Rajinder Singh, Prithvipal Singh, Morphological evaluation and sentiment analysis of Punjabi text using deep learning classification, J. King Saud University – Comput. Inform. Sci. (2018).

[3] G. Vinodhini, R.M. Chandrasekaran, A comparative performance evaluation of neural network based approach for sentiment classification of online reviews, J. King Saud University – Comput. Inform. Sci. 28 (1) (January 2016) 2–12.

[4] S. Bag, M.K. Tiwari, T.S. Chan Felix, Predicting the consumer's purchase intention of durable goods: an attribute-level analysis, J. Business Res. (2017).

[5] Z. Feng, Jianxin R Jiao, Jessie Yang, L. Baiying, Augmenting feature model through customer preference mining by hybrid sentiment analysis, Expert Syst. Appl. 89 (2017) 306–317.

[6] H. Qiongxia, X. Zheng, Z. Dong, R. Chen, Deep Sentiment Representation Based on CNN and LSTM. 2017 International Conference on Green Informatics.

[7] K. Sailaja, D. Evangelin Geetha, T.V. Sai Manoj. Analysing the Data from Twitter using R. Department of Computer Applications, MSRIT, ICRITCSA M S Ramaiah Institute of Technology, Bangalore, vol. 5, Special Issue 2, 2016.

[8] F.M.F. Wong, C.W. Tan, S. Sen, M. Chiang, Quantifying political leaning from tweets, retweets, and retweeters, Trans. Knowl. Data Eng. (2016).

[9] L. Flekov, O. Ferschk, I. Gurevych, UKPDIPF A Lexical Semantic Approach to Sentiment Polarity Prediction in Twitter Data, in: Proceedings of the 8th International Workshop on Semantic Evaluation Dublin, Ireland, 2014, pp. 704–710, 23–24.

[10] S. Bird, E. Klein, E. Loper, Natural Language Processing with Python, O'Reilly Media Inc., 2009.

[11] L. Williams, C. Bannister, M. Arribas-Ayllon, A. Preece, Spasic Lowri, The role of idioms in sentiment analysis, Expert Syst. Appl. 42 (2015).

[12] O. Kolchyna, T.P. Tharsis, P. Souza, T. Aste Treleaven, Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination, Department of Computer Science, UCL, Gower Street, London, UK, 2015.

[13] J. Serrano, P. Francisco, E. Herrera-Viedma, Sentiment analysis: a review and comparative analysis of web services, Inform. Sci. 311 (2015) 18–38.

[14] O. Araque, I. Corcuera, J. FernandoSnchez-Rada, A. Carlos Iglesias, Enhancing deep learning sentiment analysis with ensemble techniques in social applications, Expert Syst. Appl. 77 (2017).

[15] Jos M. Chenlo, David E. Losada, An empirical study of sentence features for subjectivity and polarity classification, Inform. Sci. (2014).

[16] H. Tang, Cheng X SongboTan, A survey on sentiment detection of reviews, Expert Syst. Appl. (2009) 36.

[17] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inform. Retrieval 2 (2008) 1–135.

[18] B.J. Jansen, M. Zhang, K. Sobel, A. Chowdury, Twitter power: Tweets as electronic word of mouth, J. Am. Soc. Inform. Sci. Technol. 60 (11) (2009) 2169–2188.

[19] H. Saif, H. Yulan, H. Alani, Semantic sentiment analysis of twitter, The Semantic Web-ISWC, Springer, 2012, pp. 508–524.

[20] E. Kouloumpis, T. Wilson, J. Moore, Twitter sentiment analysis: the good the bad and the omg!, ICWSM 11 (2011) 538–541.

[21] A. Montejo Rez, E. Martnez-Cmara, M. Teresa, M. Valdivia, L. Alfonso, U. Lpez, Ranked WordNet graph for sentiment polarity classification in Twitter, Comput. Speech Language (2014).

[22] Y. Han, Kim K. Ko, Sentiment analysis on social media using morphological sentence pattern model, Softw. Eng. Res., Manage. Appl., Stud. Comput. Intell., SCI, Springer 654 (2016) 85–101.

[23] L. Xiaodong, G. Jianfeng, H. Xiaodong, Deng Li, D. Kevin, Ye Yi Wang, Representation learning using multi-task deep neural networks for semantic classification and information retrieval, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 912–921.

[24] Y. Rao, O. Li, X. Mao, W. Liu, Sentiment topic models for social emotion mining, Inf. Sci. (2014).

[25] Jos M. Chenlo, David E. Losada, An empirical study of sentence features for subjectivity and polarity classification, Inf. Sci. (2014).

[26] L. Bing, Sentiment analysis and subjectivity, Second edition., An chapter in Handbook of Natural Language Processing, 2010.

[27] P. Melville, W. Gryc, R.D. Lawrence, Sentiment analysis of blogs by combining lexical knowledge with text classification, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 1275–1284.

[28] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data, Proc. Assoc. Comput. Linguistics (2011) 30–38.

[29] T. Wilson, J. Wiebe, P. Hoffman, Recognizing contextual polarity in phrase-level sentiment analysis, Proceedings of HLT/EMNLP, 2005.

[30] K. Mouthami, K. Nirmala Devi, Murali Bhaskaran, Sentiment analysis and classification based on textual reviews. Published, Information Communication and Embedded Systems, vol, (ICICES), IEEE, 2013.

[31] Y. Kim Convolutional neural networks for sentence classification. arXiv preprint arXiv: 1408.5882, 2014.

[32] W. Medhat, Korashy H. HassaAn, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113.

[33] H. Kang, S. Joo, N. Yoo, D. Han, Senti-lexicon and improved Nave Bayes algorithms for sentiment analysis of restaurant reviews, Expert Syst. Appl. 39 (5) (2012) 6000–6010.

[34] P. Soujanya, C. Erik, G. Alexander, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. volume (2016) 42–49.

[35] E. Cambria, A. Hussain, C. Havasi, C. Eckl, Sentic Computing: Exploitation of Common Sense for the Development of Emotion-Sensitive Systems, Berlin Heidelberg: Springer-Verlag. (2012), vol. 5967 of LNCS. 148-156.

[36] R. Bhayani, L. Huang. Twitter sentiment classification using distant supervision Research Gate, January 2009.

[37] S. Kim, E. Hovy, Determining the sentiment of opinions, Proceedings of the 20th international conference on Computational Linguistics, Association for Computational Linguistics, 2004, p. 1367.

[38] B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: D. Scott (Ed.), Proc. of the ACL. Morristown, 2004, pp. 271–278.

[39] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Seattle, 2004, pp. 168–177.

[40] A.-M. Popescu, O. Etzioni, Extracting product features and opinions from reviews, Proceedings of EMNLP-2005, 2005, pp. 3–28.

[41] S. Blair-Goldensohn, K. Hannan, R. McDonald, Neylon Building a sentiment summarizer for local service review, in: Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era, 2008, pp. 14–23.

[42] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, Red Opal: Product-feature scoring from review, in: Proceedings of the 8th ACM Conference on Electronic Commerce, ACM, 2007, pp. 182–191.

[43] T. Wang, Y. Cai, H. Leung, R. Lau, Product aspect extraction su- pervised with online domain knowledge, Knowl.-Based Syst. 71 (2014) 86–100.

[44] R. Collobert, J. Weston, L. Bottou, M. Karlen, language processing (almost) from scratch, J. Mach. Learn. Res. 12 (2011) 2493–2537.

[45] http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools.

[46] B. Farah, C. Carmine, P. Antonio, R. Diego, V.S. Subrahmanian, Sentiment Analysis: Adjectives and Adverbs are Better than Adjectives Alone, Published in ICWSM (2007).

[47] S. Dario, S. Gjorgji, M. Gjorgji, D. Ivica. Twitter Sentiment Analysis Using Deep Convolutional Neural Network published at: https://www.researchgate.net/publication/279208470June2015.

[48] P. Soujanya, C. Erik, K. Lun-Wei, G. Chen, G. Alexander, A Rule-Based Approach to Aspect Extraction from Product Reviews, AAAI, Quebec city, 2014, pp. 1515– 1521.

[49] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37 (1) (2011) 9–27.

[50] https://nlp.stanford.edu/sentiment/treebank.html.

Corresponding author
Paramita Ray can be contacted at: rayparamita@yahoo.com